International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249–8958, Volume-1 Issue-3, February 2012

The Design and Implementation of KDD System for Industrial Flow Object

Ming Cai, Jing Cai, Shouning Qu

Abstract**: *KDD is an important research and application area. This paper is aimed at the application of flow object's association rules extraction and object modeling in the cement industry. We adopt the improved Apriori algorithm and the flexible neural tree model of the structure optimization algorithm, designing and implementing the KDD system for industrial flow object by J2EE. The whole system is mainly divided into two functions: one function module is association rules extraction, the other one is object modeling, and the original data were collected from the decomposing furnace production link, which is one of the most important processes of the cement industry.***

Index Terms**: *Association Rule, Flow Object, J2EE, KDD, Object Modeling***

I. Introduction

KDD (knowledge discovery in databases) has been successfully applied to a number of scientific and engineering fields in recent years, i.e., image processing, expert system, machine learning, artificial intelligence and so on. Theory and technology of KDD used to solve engineering problems has become a primary task of its research. KDD is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [1],[2]. KDD makes the information into knowledge, and discovery the hidden knowledge gold from the data mine. At the same time, a large amount of depositional data is saved in the system of process industry computer control, which contains much unknown rules and knowledge. So, it has an important practical meaning that KDD technologies--Association Rules Extraction and Object Modeling are used in the industrial flow object to discovered knowledge in the mass depositional data. In addition, the flow industry is a very complex lager industrial system with the harsh production conditions and environment, i.e., high temperature, high voltage, vacuum, inflammable and explosive, what's more, each production processes is real-time, holistic and interacting[3],[4]. Obviously, the method of single-process control and human experience is unable to meet the production requirements, so we must put KDD into practice from the angle of system.

The KDD system of Industrial Flow Object which we research and try to develop is divided into two functions: Association Rules Extraction and Object Modeling. The first function module is association rules extraction in which we adopt an improved Apriori[5] algorithm-- reducing the transaction data. The other function module is object modeling in which we adopt the PIPE (Probability Incremental Program Evolution) algorithm and SA (Simulation Annealing) algorithm to build and optimize the FNT (Flexible Neural Tree) [6],[10] model. Encoding the entire system we use Java language and JSP (Java Servelet Pages), Servlet, JavaBean, Javascript, JfreeChart and some other J2EE techniques.

As has been proven in practice, the cement industry accumulated a large amount of depositional data and we can excavate association rules and object modeling from these mass depositional data by our KDD system. According to the excavated association rules we can find the influence relations between parameters of the decomposing furnace, thus, we can put forward the optimized proposals for the auxiliary decision-making of the production of decomposing furnace; According to object modeling we can quickly build the FNT model of the production process of decomposing furnace, which achieves the automatic filtering of the parameters. Thus we can calculate the decomposing furnace exit temperature predictive values with the model's mathematical formula. When get the calculated predictive value, we can take it to compare with the actual temperature of decomposing furnace exit. And after comparing we can stabilize production, optimize control. Because the process of decomposing furnace is one of the most important processes of the cement industry, the two functions of our system will play an important role in enhancing production efficiency of the cement industry.

II. KEY ALGORITHMS AND TECHNIQUES

A. Data Pretreatmen

The flow object is a complex process of huge amount of calculation, which has the diverse information model, the multivariable and close-coupling characteristic [7]. So all the collected raw data inevitably has noise and other features, even is incomplete, inconsistent. Data pretreatment is an important step in the

Manuscript published on 28 February 2012.
***** Correspondence Author (s)
Ming Cai*, School of Information Science and Engineering, University of Jinan, Jinan, China, 13964077560, (e-mail: wscmwpshh@126.com).
Jing Cai, School of Information Science and Engineering, University of Jinan, Jinan, China, 0531-82765933, (e-mail: ise_caij@ujn.edu.cn).
Shouning Qu, The Center of Information and Network, University of Jinan, Jinan, China, 0531-82765933, (e-mail: qsn@ujn.edu.cn).

KDD's process, especially data mining for the incomplete, inconsistent data. If we directly use the raw data to discovery unknown knowledge, it will reduce the efficiency of data mining, even it will lead to a wrong result. In this paper, we will introduce the methods of data cleaning and data pretreatment with a brief text. In these methods, we take full account of the features of flow object data, and have solved the problems of time-series match, data discretization and normalization.

The methods of data pretreatment are followings:

1) Time-series Match. In the flow object of cement industry, the collected data from different points restrict mutually. As a result of the continuity of process, it causes the sampling values at the same sampling time are not synchronized. So we need the time-series match. We use the way of displacement to resolve the problem: according to the lag time of each parameters to move part of data. The decomposing furnace data set after time-series matching which used to excavate association rules is Table.1. The decomposing furnace data set after time-series matching which used to object modeling is Table.2.

Table.1

id	x1	x2	x3	x4	x5
1	181.8645	0.017	7.061	760.9314	891.7897
2	179.2626	0.017	7.0652	760.9301	891.7282
3	177.356	0.017	7.0386	760.6956	891.5647
4	178.4726	0.017	7.0247	760.5084	891.4942
5	176.2289	0.017	7.0117	760.3691	891.1778
6	184.0028	0.017	7.0117	760.4307	890.8526
7	182.0856	0.017	6.9978	757.7412	890.3011
8	180.8427	0.017	7.0702	756.9321	889.8656
9	180.8427	0.017	7.061	756.5698	889.2382
10	182.2015	0.017	7.0433	756.4568	888.8771
..	..	..	..	..	..
301	179.2626	0.0181	7.0837	740.8118	889.85835

x1(t/h): Amount of raw meal; x2(%): Content of C1 exit carbon monoxide; x3(t/h): Amount of pulverized coal injection; x4(℃): Temperature of tertiary-air; x5(℃): Temperature of decomposing furnace exit

Table.2

id	x1	x2	x3	x4	x5	x6	x0
1	181.8645	0.017	7.061	760.9314	496.6204	907.5842	891.7897
2	179.2626	0.017	7.0652	760.9301	509.7875	907.198	891.7282
3	177.356	0.017	7.0386	760.6956	495.172	907.198	891.5647
4	178.4726	0.017	7.0247	760.5084	508.3392	907.198	891.4942
5	176.2289	0.017	7.0117	760.3691	498.9466	907.198	891.1778
6	184.0028	0.017	7.0117	760.4307	533.7517	907.198	890.8526
7	182.0856	0.017	6.9978	757.7412	491.4413	907.198	890.3011
8	180.8427	0.017	7.0702	756.9321	475.4213	907.198	889.8656
9	180.8427	0.017	7.061	756.5698	511.192	907.198	889.2382
10	182.2015	0.017	7.0433	756.4568	534.6734	907.198	888.8771
..	..	..	..	..	..	..	..
300	176.3448	0.0181	7.0702	735.3049	515.8444	907.198	890.27465

x1(t/h): Amount of raw meal; x2(%): Content of C1 exit carbon monoxide;x3(t/h): Amount of decomposing furnace pulverized coal;x4(℃): Temperature of tertiary-air;x5(A): Main motor current of kiln ;x6(r/m): Rotational speed of high-temperature fan;x0(℃): Temperature of decomposing furnace exit

2) Data Discretization. The data discretized is used to excavate association rules. At first, the association rule algorithm was aimed at the Boolean attribute in the Large-scale commercial database. How to discretize the quantitative attributes in order to better analyze the mass raw data in the flow industry database is a primary problem of data mining applied to control engineering fields. We need to research the relationship between the important parameters of cement flow industry, so we discretize the data of table-1 into three values: '-1', '0', '1+'. The '-1' means drop. The '0'means unchanged. The '1+' means rise. The discretized data set is Table.3.

Table.3

id	x1	x2	x3	x4	x5
1	-1	0	1+	-1	-1
2	-1	0	-1	-1	-1
3	1+	0	-1	-1	1+
4	-1	0	-1	-1	-1
5	1	0	0	1+	-1
6	-1	0	-1	-1	-1
7	-1	0	1+	1+	-1
8	0	0	-1	-1	-1
9	1+	0	-1	-1	-1
10	-1	0	-1	1+	-1
..	..	..	..	..	..
300	-1	0	1+	1+	-1

3) Data Normalization . In the function module of object modeling, the value of data must between 0 and 1, so we need normalize the data of Table.2. The normalization formula is (1):

$$Y = {(X - MIN)}\Big/{(MAX - MIN)} \qquad (1)$$

The normalized data set is Table.4.

Table.4

id	x1	x2	x3	x4	x5	x6	x0
1	0.5658	0.017	0.4594	0.6094	0.3518	0.8098	0.7557
2	0.3767	0.017	0.4747	0.6094	0.4138	0.7904	0.7532
3	0.2381	0.017	0.3780	0.6055	0.34504	0.7904	0.7464
4	0.3192	0.017	0.3275	0.6024	0.40702	0.7904	0.7435
5	0.1562	0.017	0.2802	0.6001	0.3628	0.7904	0.7304
6	0.72128	0.017	0.2802	0.6011	0.5266	0.7904	0.71696
7	0.5819	0.017	0.2297	0.5566	0.3274	0.7904	0.6941
8	0.4915	0.017	0.4929	0.5433	0.2521	0.7904	0.6761
9	0.4915	0.017	0.4594	0.53731	0.4204	0.7904	0.6501
10	0.5903	0.017	0.3951	0.5354	0.5309	0.7904	0.6351
..	..	..	..	..	..	..	..
300	0.1646	0.0181	0.4929	0.1856	0.4423	0.7904	0.6930

B. *Improvement of Apriori Algorithm*

In the function module of association rules extraction, the improved Apriori algorithm is a method of reducing transaction data. The main supposition is that if a transaction data can't support any frequent set of K-item, it must not support any frequent set of K+1-item. The method scan the data to get the K-item, at the same time it will add a mark of 'bool' whose starting value is 'false' for every transaction data. In the process of counting the times of appearance of candidate's items of k, if a transaction data don't contain any candidate's items of k, its value of 'bool' will changed to 'true', or its value of 'bool' is 'false', and the value of 'bool' will not changed. If the value of 'bool' of a transaction data finally is 'true', it means that the transaction data can be deleted. Thus, the time of the next scan will reduce and the efficiency of the algorithm is improved.[8]

The discretized data set after add the mark of 'bool' is Table.5:

Table.5

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249–8958, Volume-1 Issue-3, February 2012

id	x1	x2	x3	x4	x5	bool
1	-1	0	1	-1	-1	false
2	-1	0	-1	-1	-1	false
3	1	0	-1	-1	1	false
4	-1	0	-1	-1	-1	false
5	1	0	0	1	-1	false
6	-1	0	-1	-1	-1	false
7	-1	0	1	1	-1	false
8	0	0	-1	-1	-1	false
9	1	0	-1	-1	-1	false
10	-1	0	-1	1	-1	false
..	..	..	..	..	..	..
300	-1	0	1	1	-1	false

C. Build and Optimization of Flexible Neural Tree

In the function module of object modeling, the algorithm we adopt is from the references [6],[14] -- structure optimization algorithm based on Flexible Neural Trees. In the process of the build and optimization of Flexible Neural Tree, PIPE [11],[12] algorithm realizes the structure optimization of FNT [10] model and SA[13] algorithm realizes the parameters optimization.

The entire description of the algorithm is as follows:

Step 1: Set the initial values of PIPE and SA algorithm parameters. The best individual is null and its fitness value is a very big real number (0.99999) and the value of mean error rate is ER=0.003 and the initial values of evolutional generation is 0.

Step 2: Create initial population: flexible neural trees and their corresponding parameters.

Step 3: Optimize FNT's structure by PIPE and compute fitness values and mean error rates using fitness function and the minimum of mean error rates in this generation is E.

Step 4: If the minimum value E is less than ER, then go to step 5; otherwise go to step 2.

Step 5: The optimum tree structure is found, then optimize its parameters using SA, and all the parameters used in the best tree formulated a parameter vector that are optimized by local search, and the structure or architecture of flexible neural tree model is fixed.

Step 6: If the maximum number of iterations of SA algorithm is reached, or no better parameter vector is found for a significantly long time (100 times) then go to step 7; otherwise go to step 5.

Step 7: The optimum model is formed, and then loop termination.[14]

D. Techniques of Javabean, Servlet, Javascript, JfreeChart

JavaBean is a reusable component written by the JAVA language, which mainly takes charge of the transactions processing, such as data operation. Using JavaBean to package the transaction logic can better realize the separation of transaction logic and page, so that the system has a better toughness and flexibility. One of the most attractive places of the JSP is that it can combine JavaBean to extend the function of pages. The mainly advantages of using JavaBean is as follows: (1) Separation of logic and page. JSP only need to take charge of the showing of page and no need to consider the background logic and the complex algorithm. (2) Reusability. It is convenient for extending the function of system. (3) Simplify the page development, convenient for modifying. If we embedded the complex algorithm into the JSP, it would not be easy to modify and maintain the pages.

Servlet is a kind of Java application procedure in Server, which is used for controlling pages transfer. Compared with the CGI (Common Gateway Interface) script, Servlet has the advantages of high efficiency, low spending, easy to development, powerful functions, portability and so on. The system of KDD of we develop is based on the mode of JSP+JavaBean+Servlet. We use the high performance Server procedure of Servlet as the total control procedure which plays a role of control. Under the request of HTTP, Servlets implement interactively browsing and adding data, then call different JSP programs and then use JSP tags or Javascript to generate dynamic content on the page.

Javascript is a client script language based on object and event-driven, and it is also relative safety. Javascript is often used to add dynamic functions for HTML page, such as response each kind of user's operations. Moreover, Javascript also has the powerful drawing functions. In the second function module of our system, achieving the drawing of FNT model is by the mxGraph drawing components of Javascript.

JFreeChart is an open charting class library on the JAVA platform. It is entirely written by the JAVA language, and designed for the usage of Servlets, JSP, and so on. JFreeChart can generate pie charts, bar charts, line charts, etc.. In the second function module of our system, generating a trend chart is realized just by the JfreeChart.

III. OVERALL DESIGN AND IMPLEMENTATION OF THE SYSTEM

A. Data Flow Chart

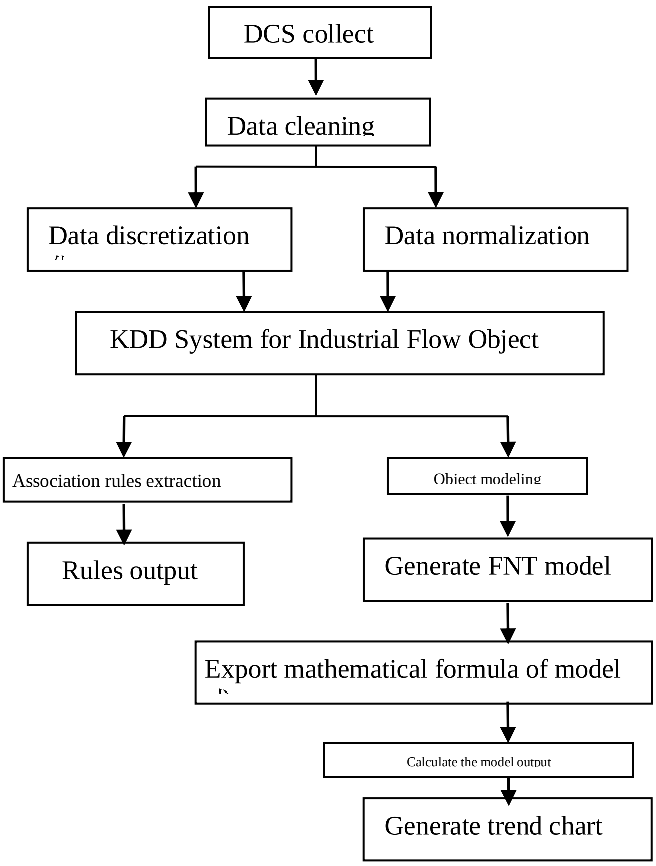

Fig.1 Data flow chart

B. Structure of the System

The entire system mostly divided into two functions. The chart of overall function structure is as follows:

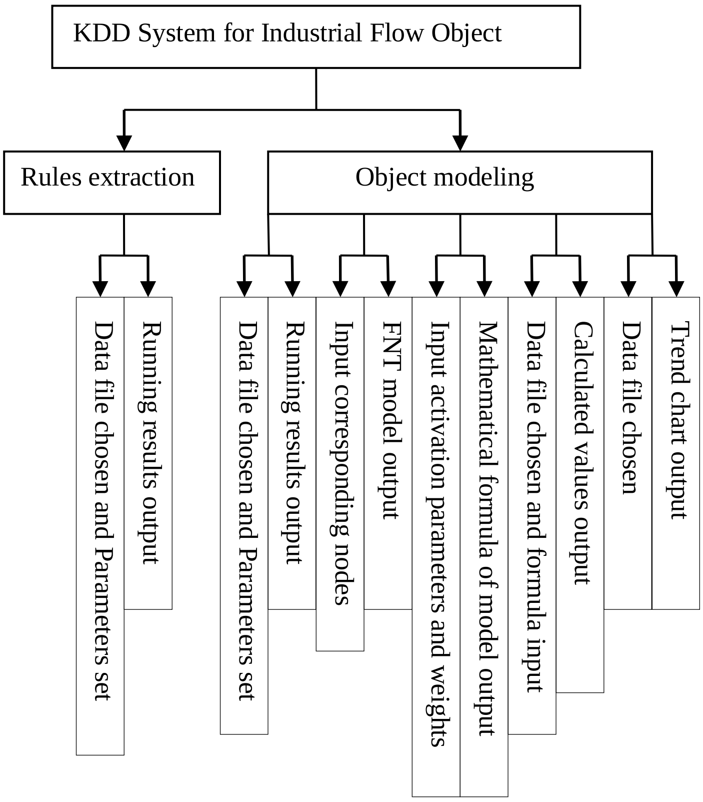

Fig.2 Structure of the system

The first module only has one function that is association rules extraction. The second module has five functions that are object modeling, generate FNT model, export mathematical formula of model, calculate the model output, generate trend chart.

The home page of the system is as follow:

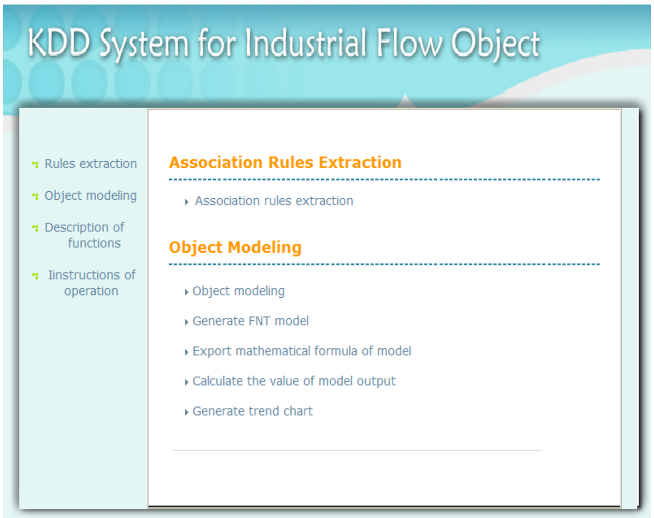

Fig. 3 Home page

C. Description and Implementation of the System's Modules

1. Association Rules Extraction.

In the page of association rules extraction, choose the treated decomposing furnace data file (namely Table-5), input the values of minimum support and minimum confidence which both range from 0 to 1, press the OK button, then we will dig out the rules of the five parameters in decomposing furnace. These rules' confidence all bigger than the minimum confidence set before. Through the analysis of extracted rules, we can find the influence relationship between some parameters fluctuation and other parameters changes. As well as we can get some optimization suggestions by summing up these rules, for examples, the rule 17(amount of raw meal Drop, amount of pulverized coal injection Rise ==> temperature of decomposing furnace exit Drop), this indicates that when the amount of raw meal reduce and the amount of pulverized coal injection increase, the temperature of decomposing furnace exit will not be rise, but is drop. The reason is that when the amount of pulverized coal is injected too much, the incompletely burnt or unburnt pulverized coal will enter into the last-level preheater and continue to burn, which makes the bottom of preheater's temperature higher than the temperature of decomposing furnace exit. At the same time, it is also easy to lead to the yield and quality of clinkers declining, because the preheater's crust is choked and the temperature of decomposing furnace is not high enough. Thus it can be seen that the control of the amount of raw meal and the amount of pulverized coal injection is so important for the yield and quality of clinkers.

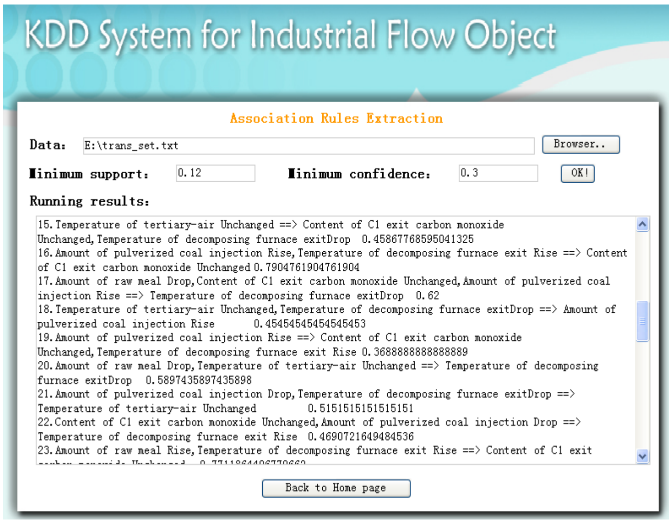

Fig.4 Page of Association rules extraction

In this function module, the key algorithm is the improved Apriori algorithm which has been introduced in detail at the section 2 of chapter 2, and the key technique of J2EE is the algorithm encapsulation by JavaBean. Adding the compiled algorithm to a bean-bag as a independent class, we can solely modify the algorithm and the page. Here, Servlet takes charge of data transfer. Servlet transmits the data inputted in the page to the JavaBean, then JavaBean starts to operate and packs the operation result, and then the result is accessed by JSP procedure.

2. Object Modeling

The second function module as introduced before is divided into five sub-functions: object modeling, generate FNT model, export mathematical formula of model, calculate the model output, generate trend chart.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

◆ Object modeling

In the page of object modeling, choose the treated decomposing furnace data file (namely Table-4), and set the values of FNT model parameters. The number of terminal instruction set is '6' means that there are six terminal instructions (because we have six parameters:x1: Amount of raw meal; x2: Content of C1 exit carbon monoxide; x3: Amount of decomposing furnace pulverized coal; x4: Temperature of tertiary-air; x5: Main motor current of kiln; x6: Rotational speed of high-temperature fan); The number of function instruction set is '3' means that there are tree function instructions. The tree function instructions respectively are'+2','+3','+4' ('+2' means there are two input flexible neuron, by this analogy..). The minimum depth and maximum depth respectively means the values of FNT's depth. Here, we set the value of '2' and '3'. After all values been set, press the OK button, we can get the results: the last generation of the best individual nodes, the parameter of activation function, and the weights of FNT's branches. These results are prepared for the behind four sub-functions.

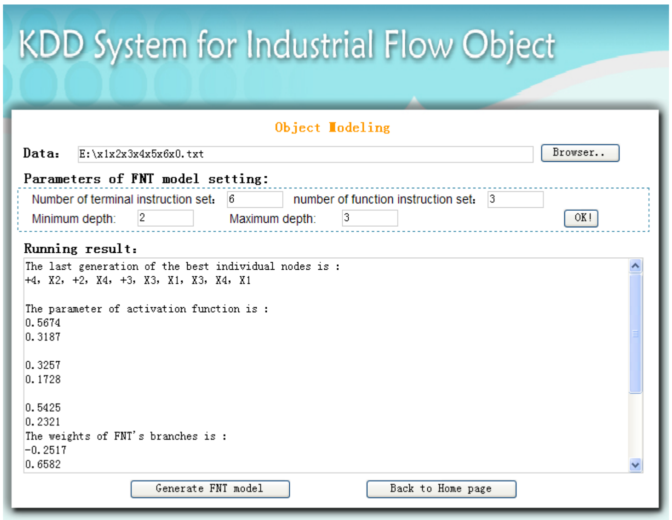

Fig. 5 Object modeling Page

In this sub-function module, the key algorithm is from the reference [6,14] -- structure optimization algorithm based on Flexible Neural Trees, which has been introduced in detail at the section 3 of chapter 2, and the key technique of J2EE is also the algorithm encapsulation by JavaBean.

◆ Generate FNT model

In the page of generating FNT model, input the last generation of the best individual nodes which got from the first sub-function, press the OK button, the page will generate the FNT model. This model's output parameter is chosen in advance. In our system we chose the temperature of decomposing furnace exit as the model's output parameter. According to the FNT generated on the page, we can find that influencing the temperature of decomposing furnace exit are the parameters of x1(Amount of raw meal), x2(Content of C1 exit carbon monoxide), x3(Amount of decomposing furnace pulverized coal), x4(Temperature of tertiary-air). This result is as the same as the sum-up of workers' experience and the analysis at the decomposing furnace structure and internal burning mechanism [15].

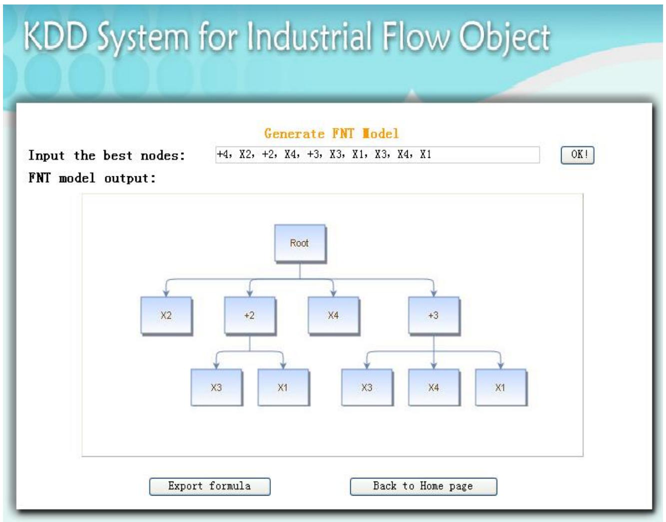

Fig. 6 Generate FNT model Page

To generate the FNT model in term of the input nodes, the key technique is that call the 'mxclient.js' control from the mxGraph drawing components.

◆ Export mathematical formula of model

In the page of export mathematical formula of model, input the parameters of activation function, and the weights of FNT's branches, press the OK button, we will get model's corresponding mathematical formula. The method of calculation is from the reference [14]--flexible activation function. The output of a flexible neuron can be calculated as follows:

$$out_n = f(a_n, b_n, net_n) = e^{-\left(\frac{net_n - a_n}{b_n}\right)^2} \quad (2)$$

$$net_n = \sum_{j=1}^{n} w_j * x_j \quad (3)$$

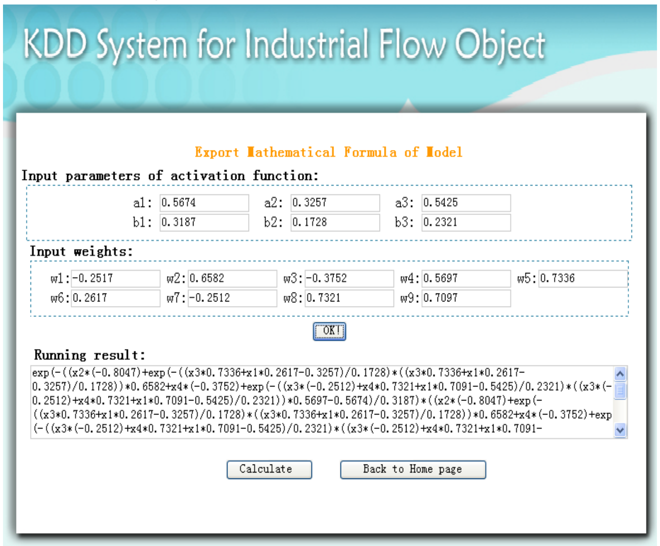

Fig. 7 Page of Export mathematical formula of model

To realize the function, the key is to get the parameters of activation function and the weights of FNT's branches through object modeling, then use the series of

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

mathematical formulas to code.

◆ Calculate the value of model output

In the page of calculating the value of model output, choose the data file and input the mathematical formula of model which generated at the front page, press the OK button, we will get the values of the model output.

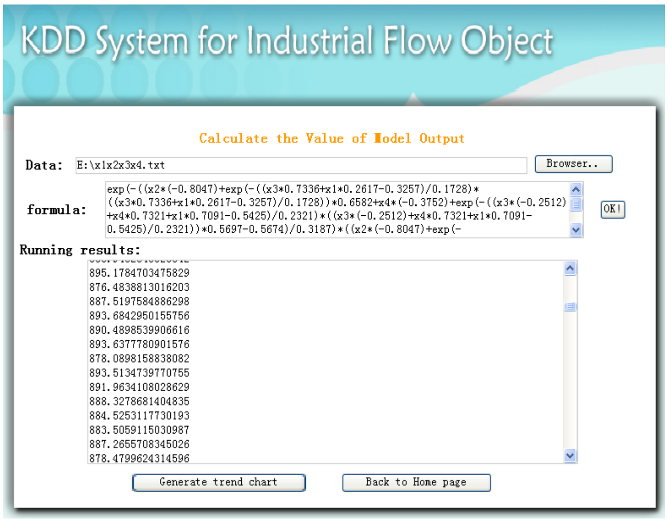

Fig. 8 Page of Calculate the model output

◆ Generate trend chart

In the page of generating trend chart, choose the data file, press the OK button, we will get a trend chart about the contrast between the model output and the actual temperature of decomposing furnace exit. Using the trend chart, we can more intuitively find the difference.

This function is that according to the imported data file can generate a trend chart. To develop this function we just need to import the relational Jfreechart controls into the application.

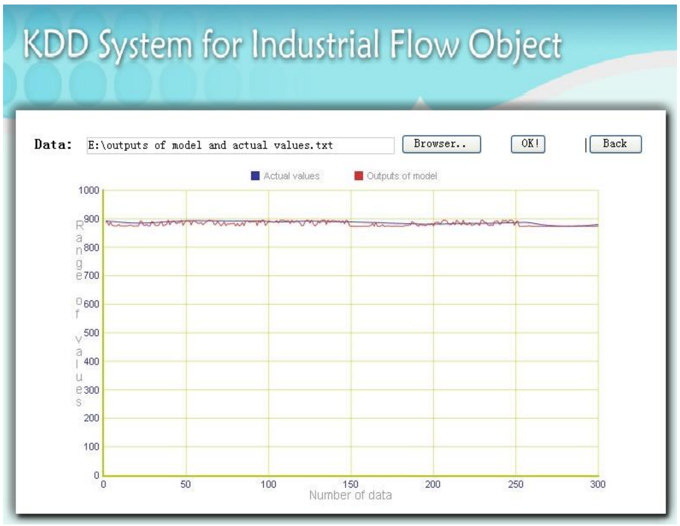

Fig.9 Generate trend chart

IV. Conclusion

The design and realization of the system of KDD for Industrial Flow Object is on the basis of the existing theories--improved Apriori algorithm and the algorithm of build and optimization of Flexible Neural Tree. What we research and develop is to realize the association rules extraction and object modeling from the angle of system, and perform it with dynamic Web page forms by the techniques of J2EE. It is our goal that applies the final solution to the actual production control of the cement which is controlled by DCS (distributed control system). With this application the DCS can reach the goal of improving the cement's production and quality, stabilizing the production line parameters, and finally improving the benefits of cement enterprise.

References

1. Mitra S., Pal S., Data Mining in Soft Computing Framework: A Survey, IEEE Trans on NN,2002,13(1):3-13.
2. Johnson J., Liu M., Unification of knowledge discovery and data mining using rough sets approach in a real-world application. RSCTC 2002,LNAI 2005,2001:330-337.
3. R. K. Xu, J. m. Wang, Q. R. Jiang, Application Status of cement production automatization technology and equipment in China, China Building Material, no.05,1999,pp. 1-5.
4. J. L. Li, H. Q. Zhou, The Production of Cement, Wuhan University of Technology Press, Wuhan, 2008.
5. Agrawal R, Imielinski T, Swami A, Mining association rules between sets of items in large database, Proceedings of the ACM S IGMOD Conference on Management of Data, Washington D. C, 1993. 207-216.

A. *F. Fu, Study and Implementation of Process Object Modeling, M. S. Thesis, Department of Information Science and Engineering, University of Jinan, June 2010.*

6. S. N. Qu, Study on the Knowledge Extraction Method for Flow Object and its Application in Flow Industrial Control, Ph. D. Thesis, Beijing Institute of Technology, Beijing, May 2010.
7. G. Q. Qiang, Analyzing and Optimizing of Industrial Parameters Based on Data Engineering, M. S. Thesis, Department of Information Science and Engineering, University of Jinan, May 2009.
8. Z. L. Liu, Analyzing and Optimizing of Decomposing Furnace Parameters Based on History Data M. S. Thesis, Department of Information Science and Engineering, University of Jinan, May 2009.
9. Y. H. Chen, B. Yang, J. W. Dong et al, Time-series Forecasting Using Flexible Neural Tree Model, Information Science, 2005, 174: 219-235.
10. R. P. Salustowicz, J. Schmidhuber, Probabilistic Incremental Program Evolution, Evol, Comput, 1997, 2 (5): 123–141
11. Y. H. Chen, S. Kawaji, System Identification and Control using Probabilistic Incremental Program Evolution Algorithm, Journal of Robotics and Mechatronics, 2000,12:(6), 657-681
12. X. Y. Cui, Y. H. Chen, K. F. Shi, B. Yang, Application of artificial neural networks-genetic algorithms to prediction of cement strength, Journal of Shandong Institute of Building Materials, 1998,(3),275-277
13. S. N. Qu, A. F. Fu, Z. L. Liu, etal.The Improvement and Application of Structure Optimization Algorithm Based on Flexible Neural Trees, In:2009 International Conference on Information Technology and Computer Science,2009
14. F. S. Wang. Basic knowledge about modern cement production, China building material industry press, Beijing, 2004

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)